RESEARCH ARTICLE

Natural genetic variation in mitochondrial health-regulating genes in the C. elegans strains CX11314 and EG4725 leads to increased mitochondrial resilience.

Aaryan Samanta[1,*], Mansi Mandal[2], Mohammed Yasar Meeran[3]

[1]Legend College Preparatory School, aaryan.samanta@gmail.com

[2]Mountain House High School, mansimandal999@gmail.com

[3]Lebanon Train High School, yasar.meeran@gmail.com

* Corresponding author email: aaryan.samanta@gmail.com

Abstract

Mitochondria play a central role in maintaining cellular function and homeostasis, and their dysfunction can severely impact lifespan and aging. Owing to their conserved pathways, human orthologs, and experimental tractability, Caenorhabditis elegans was used as the model organism in this study. We investigated natural genetic variation in the C. elegans strains CX11314 and EG4725, which are characterized by increased resistance to mitochondrial dysfunction. Specifically, we examined single nucleotide polymorphisms (SNPs) and differential expression in six key mitochondrial health–regulating genes—atfs-1, eat-3, fzo-1, pdr-1, drp-1, and pink-1—that control mitochondrial morphology and mitophagy, using variant analysis, phylogenetic comparison, and RNA-seq data. Our analyses identified strain-specific SNPs in these genes and revealed differences in their expression levels, suggesting that natural genetic variation may modulate mitochondrial resilience. Building on these findings, we propose applying CRISPR–Cas9–mediated homology-directed repair (HDR) genome editing to directly test how specific variants in mitochondrial-regulating genes give rise to distinct phenotypic and behavioral outcomes.

Keywords

C. elegans, Mitochondrial Dysfunction, Mitophagy, Mitochondrial Morphology, Mitochondrial Resilience

Introduction

According to the Cleveland Clinic, approximately 1 in every 5,000 individuals is affected by a genetic mitochondrial disease[1,2]. These diseases can affect nearly every organ, ranging from the brain and nervous system to the heart and liver, emphasizing the importance of analyzing these conditions through the various related genes and genetic pathways.

Its transparent body allows scientists to easily observe internal structures and track subtle changes in development, behavior, and gene expression with ease[3]. Its short lifespan helps researchers quickly observe the effects of genetic or environmental changes across an entire lifespan, making experiments more efficient. Additionally, since *C. elegans* are hermaphrodites, they will be able to produce offspring through self-fertilization. This minimizes genetic variation in the tests since all offspring will be genetically identical across generations. Furthermore, researchers can directly translate discoveries made with *C. elegans* to human diseases, as there are conserved genetic pathways and functional orthologs.

It is crucial to note that varying *C. elegans* strains should be used, as the majority of studies pertaining to *C. elegans* have been conducted on the N2 (Bristol) strain[4]. However, conducting research on strains with unique abilities like heteroplasmy resistance will reveal the genetic variations that result in that characteristic. Therefore, this study will focus on two *C. elegans* strains that have shown improved mitochondrial function and resilience: CX11314 and EG4725. By identifying genes and expression

pathways that contribute to their improved mitochondrial morphology and mitophagy, we aim to discover novel insights into the biological mechanisms that promote healthy aging. These discoveries will then have the potential to facilitate research into age-related decline in humans.

Previous research has found that mitochondrial dysfunction accumulates as individuals age, and removing functionally compromised mitochondria helps organisms recover from the adverse effects. In fact, research in fruit flies has shown that by removing damaged mitochondria, male and female fruit flies were able to live 20 percent and 12 percent longer, respectively. Additionally, previous studies have discovered cellular mechanisms that maintain mitochondrial health and slow down the aging process. In the N2 *C. elegans* strain, mitochondrial fission genes(*drp-1*), fusion genes(*fzo-1* and *eat-3*), mitophagy genes(*pink-1* and *pdr-1*), and mitochondrial unfolded protein response(UPR) genes(*atfs-1*) have been identified. These findings have identified that these genes assist in maintaining mitochondrial structure, responding to stress, and removing damaged mitochondria[6,7,8]. However, when these processes fail, cells accumulate dysfunctional mitochondria, contributing to aging, neurodegeneration, and other chronic diseases[9]. As mitochondrial dysfunction continues to emerge as a key factor in many human diseases, understanding the genetic control of mitochondrial maintenance is more relevant than ever. Yet, despite these advances, most research relies on lab-adapted strains and overlooks the natural genetic diversity that could reveal new, potentially therapeutic, pathways to promote mitochondrial resilience and healthy aging.

Future studies are also focusing on studying the evolutionary adaptations and genetic differences between different strains of *C. elegans*. This shift from single-strain to multi-strain comparisons can help researchers uncover many evolutionary adaptations in *C. elegans* that are caused by the different environmental pressures. In a study by Zhang et al., researchers analyzed gene expression across 207 wild *C. elegans* isolates and showed that natural diversity in gene expression and possible regulatory mechanisms highlight the promise of using gene expression variation to understand how phenotypic diversity is generated[10]. Furthermore, recent research projects have started to use natural genetic variation in *C. elegans* strains as a tool for discovering new aging regulators. For example, a paper by Yin et al. discusses how different *C. elegans* isolates show diverse lifespans and age-related declines. In the study, they identified natural genetic factors that influenced the rate of aging in *C. elegans* by looking at the different aging mechanisms among the different *C. elegans* isolates and comparing them to their genetic variance[11].

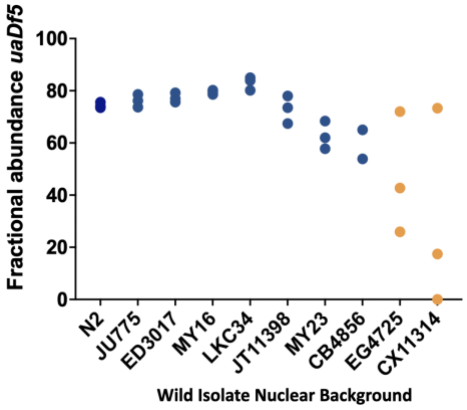

Figure 1. This figure shows the fractional abundance of the protein uaDf5 for each wild isolate strain. This figure is based on unpublished data provided by the Rothman Laboratory at UCSB.

The field of aging genetics in *C.* elegans has directed its focus on mito-nuclear interactions and post-transcriptional control under stress. Recent studies have also found that mitochondrial function and aging are regulated by these interactions. Furthermore, lifespan is a polygenic trait heavily influenced by approximately 293 nuclear regions as well as widespread mito-nuclear interactions, with worms with enhanced interactions living significantly longer[12]. In recent years, studies have been utilizing gene

expression data for analysis. Unfortunately, there remains a major gap in understanding how SNPs and varying gene expression improve mitochondrial health. Figure 2 demonstrates the key genetic pathways that are involved in the regulation of mitochondrial function and the dysfunction of mitochondria.

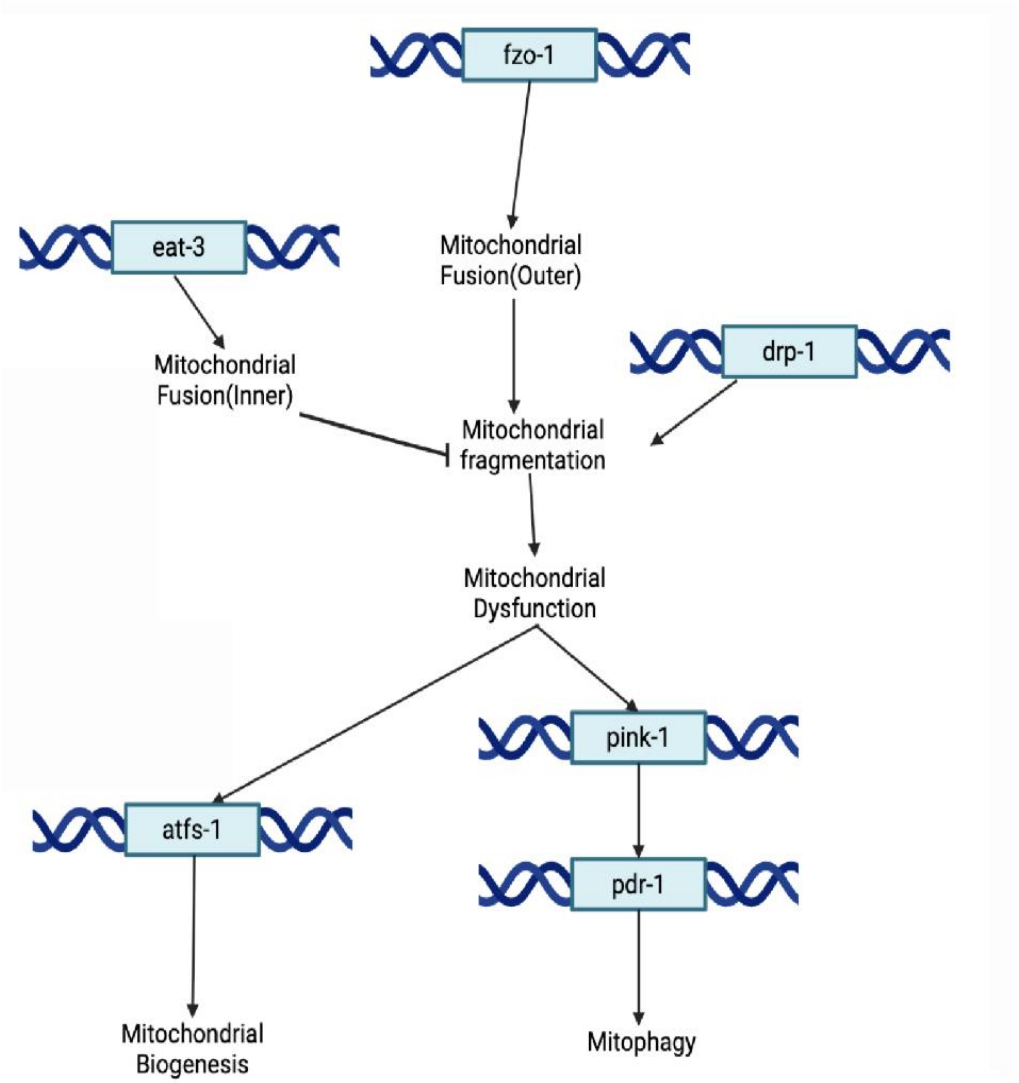

Figure 2. Overview of mitochondrial quality control pathways with mitochondrial and nuclear genes

This leads to our research question: What genetic variants in mitochondrial morphology and mitophagy-related genes contribute to the regulation of mitochondrial quality and heteroplasmy in the *C. elegans* strains CX11314 and EG4725 compared to the N2 reference strain? We hypothesize that genetic variation and variable expression in mitochondrial morphology and mitophagy-regulating genes (*drp-1, fzo-1, pink-1, eat-3, pdr-1,* and *atfs-1*) enhance mitochondrial quality control in the wild *C. elegans* strains CX11314 and EG4725, leading to reduced mitochondrial dysfunction and lower levels of heteroplasmy compared to the N2 reference strain. Our study aims to answer these questions by analyzing genomic variants and expression levels through the utilization of bioinformatics.

Methods

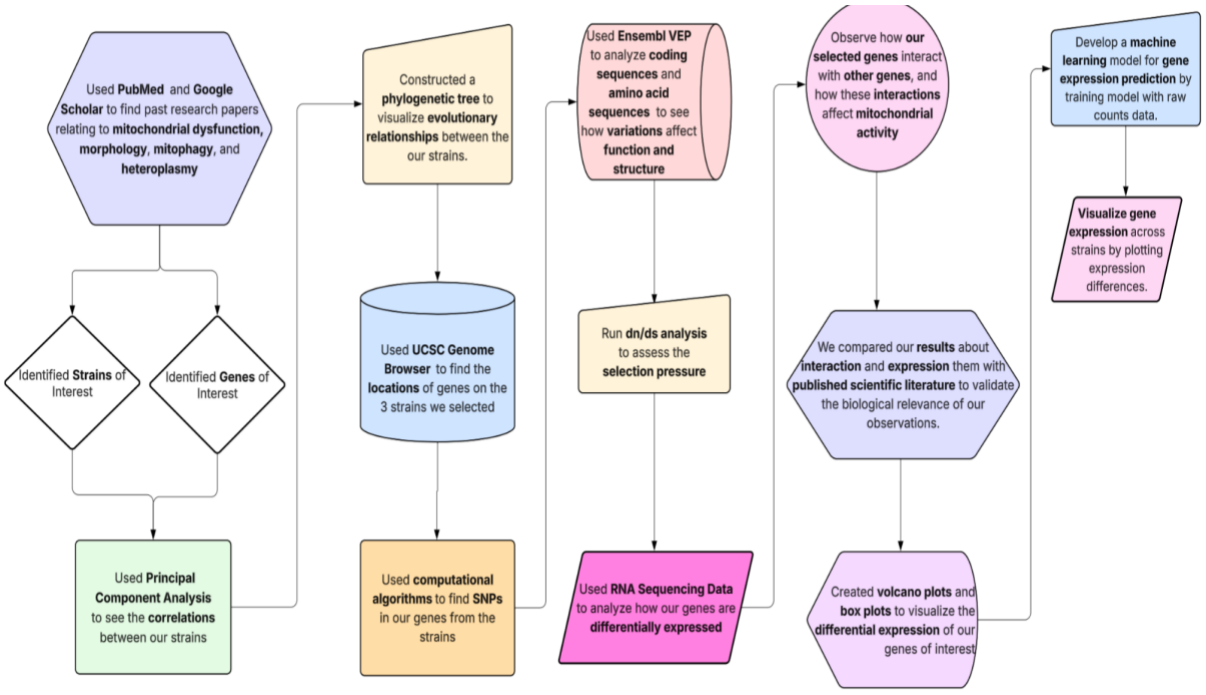

Figure 3. *This figure shows the step-by-step methodology used in this study. Diagram created using Lucidchart.*

To investigate the genetic differences underlying mitochondrial function in *C. elegans* strains, we designed a multi-step analysis combining literature review, variant analysis, phylogenetics, and differential gene expression. We began by using PubMed, PMC, EuroPMC, and Google Scholar to analyze past research papers relating to mitochondrial morphology, mitophagy, dysfunction, and heteroplasmy. From this, we selected N2, CX11314, and EG4725 for further analysis. CX11314 and EG4725 were chosen due to their reported improvements in mitochondrial function compared to the standard N2 strain. To investigate the genetic basis of these differences, we focused on genes involved in mitochondrial processes that are conserved between *C. elegans* and humans. By comparing these strains to the N2 strain, we found key differences in CX11314 and EG4725 that would enhance mitochondrial quality.

After obtaining Variant Call Format(VCF) files from CaeNDR, a *C. elegans* natural diversity resource, bioinformatics Python libraries such as scikit-allel(v1.3.13), Biopython(v1.85), ETE3(v3.1.3), PyDESeq2(v1.40.2), NumPy(v2.3.2), Matplotlib(v3.10.8), and SciPy(v1.16.0) were utilized for further analysis[13-20]. Principal Component Analysis(PCA), a statistical method commonly used to simplify complex, high-dimensional data while preserving variance, was used to compare variants between 540 isotypes. This visualization helped assess the genetic similarities, where the distance between each strain increases with genetic divergence.

A phylogenetic tree was then used to visualize the overall evolutionary relationships and genetic distances among the 12 selected *C. elegans* isotypes: MY1, AB1, EG4725, CB4856, N2, MY16, JU393, ED3017, DL238, LKC34, GXW1, and CX11314. The UCSC genome browser, an online tool that allows scientists to analyze DNA from different organisms, was used to find the locations of our selected genes in the *C. elegans* genome. The gene ranges obtained from the genome browser helped us identify genetic differences in our strains of interest by consulting the VCF data.

To identify differentially expressed genes, we processed RNA-seq counts data for DESeq2 analysis, comparing the CX11314 and N2 strains. Using the VCF data, we identified SNPs where the strains differed

and generated variant tables, helping us pinpoint mutations that could affect gene function and explain expression differences from our RNA-seq results.

The Ensembl Variant Effect Predictor (VEP) was used to compare DNA, RNA, and protein sequences to identify nucleotide and amino acid differences[21]. This helped us identify how genetic variations affect protein structure and function, thereby impacting mitochondrial health and function. To determine whether the observed genetic differences in our genes of interest are driven by natural selection, a Nonsynonymous/Synonymous Substitution Ratio (dN/dS) was calculated on our candidate genes.

Gene expression data of 609 samples from 208 different strains were attained from the NCBI GEO dataset GSE186719, which included both raw read counts, the sum of the number of times a DNA or RNA was observed in an experiment, and TPM (Transcripts per million) values for the samples, in order to investigate how key mitochondrial-related genes are differentially expressed between CX11314 vs N2 and EG4725 vs N2[22]. Transcript isoforms were consulted using WormBase, a curated database for *C. elegans*, which provides transcript names, genomic coordinates, and functional annotation, and then the corresponding expression values were extracted. Differential expression analysis was performed with the help of PyDESeq2, which models differential gene expression with a negative binomial distribution, on raw counts, while TPM values were used for clearer visualizations. It is important to note that EG4725 was not included in the GSE186719 RNA-seq dataset. To address this limitation, we selected the most genetically similar *C. elegans* strain to represent EG4725. Based on our earlier PCA analysis of 540 isotypes, NIC266 was identified as the closest strain to EG4725 in terms of genetic similarity. Additionally, VCF analysis confirmed that NIC266 has very few SNP differences relative to EG4725, supporting its use as a reliable substitute. Therefore, RNA-seq expression data from NIC266 were used in place of EG4725 for differential expression analysis and visualization. This allowed us to make biologically meaningful inferences without introducing bias into the analysis.

Following the DESeq2 analysis, we visualized differentially expressed genes using volcano plots and box plots. Volcano plots were used to highlight genes with statistically significant fold changes, using Benjamini-Hochberg(BH) adjusted p-value < 0.05, allowing us to quickly identify upregulated and downregulated genes in our strains of interest[23]. Box plots, generated using TPM data, displayed the distribution of expression levels for specific genes across N2, CX11314, and EG4725, offering a more comprehensive comparison. Due to the lack of RNA-seq expression data for EG4725, a machine Learning approach was used to predict EG4725's gene expression profile. After cleaning the data, a Neural Network coupled with a Leave One Out Cross Validation(LOOCV) was performed, which excluded the *C. elegans* strain NIC266 from the training dataset to use for validation, as it is genetically closest to EG4725. Once the model was developed using Python packages such as Keras(v3.10.0), TensorFlow(v2.19.0), Scikit-Learn(v1.7.1), and Scipy(v1.16.0), Spearman's correlation was then used to evaluate the EG4725's expression prediction with high confidence[24]. To further understand the biological implications of these varying expressions, we investigated how our genes of interest interacted with other genes to understand the functional pathways. We validated the relevance of these interactions by cross-referencing our results with existing research.

Results

CX11314 and EG4725 Are Genetically And Evolutionarily Divergent From N2.

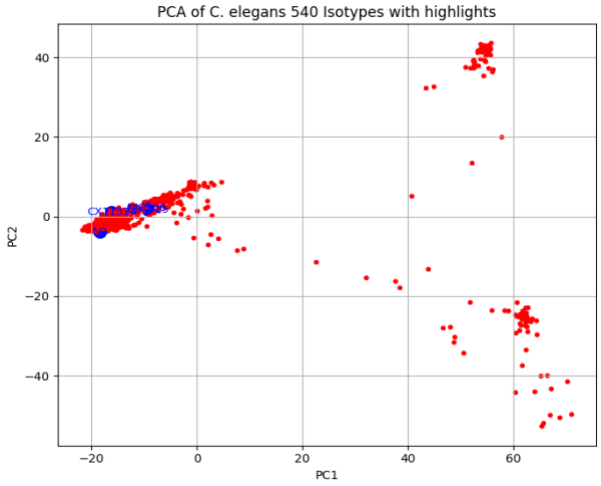

Figure 4. *Principal Component Analysis(PCA) depicting the clustering of genetically similar C. elegans strains.*

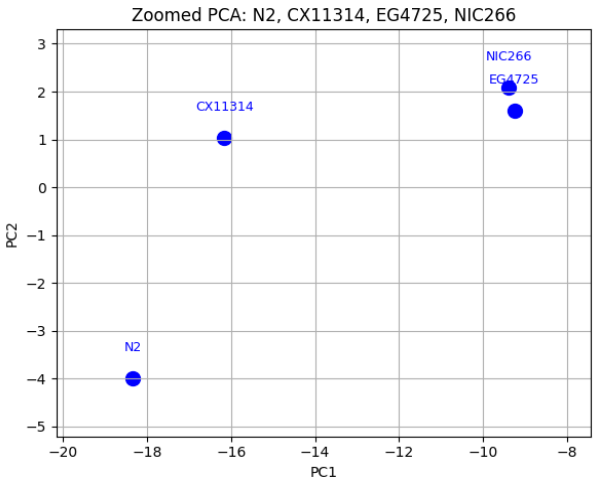

Figure 5. *A zoomed-in view depicting the C. elegans strains of interest.*

A principal component analysis (PCA) was then utilized to visualize the genetic distance between the 540 different *C. elegans* strains using the SNP data. In both Figures 4 and 5, the blue highlighted data points represent N2, CX11314, and EG4725. Additionally, as the gene expression data for EG4725 was limited, NIC266, the strain genetically closest to EG4725, was used as a proxy.

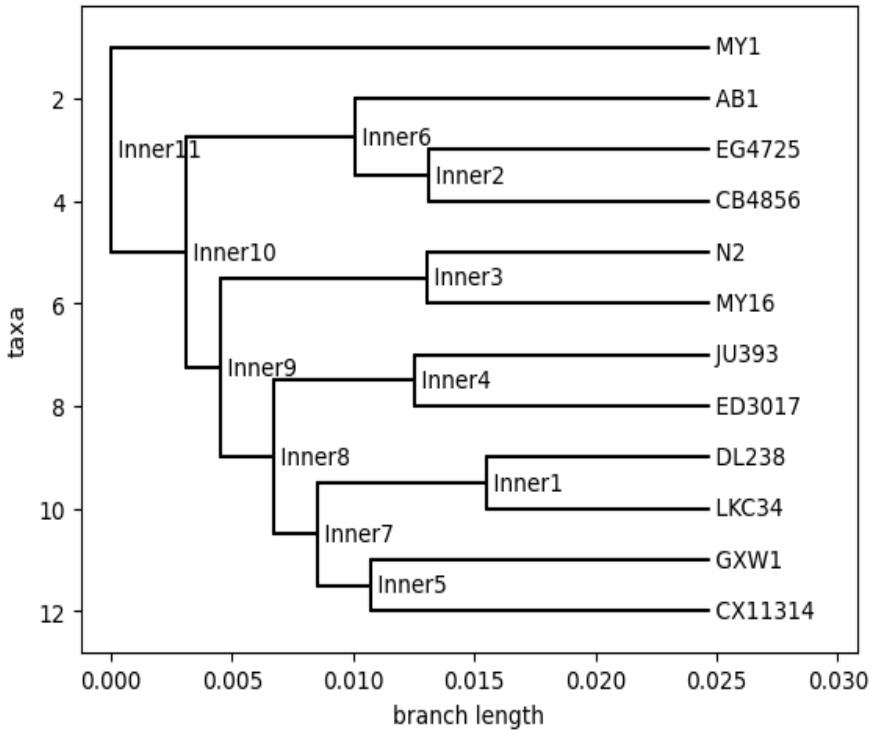

Figure 6. *Phylogenetic tree showing evolutionary relationships in different C. elegans strains.*

The application of a phylogenetic tree using SNP data also illustrated the evolutionary relationships between our strains. Figure 6 demonstrated that CX11314 and EG4725 were not only genetically divergent but also evolutionary divergent. For example, the last common ancestor between N2, CX11314, and EG4725 was at the Inner10 node, highlighting how these strains did not diverge recently.

Purifying and Neutral Selection in Genes

Table 1. Ratio of non-synonymous to synonymous mutations for respective genes.

Gene	dN	dS	dN/dS	Significance
atfs-1	34	41	0.829	Purifying
pink-1	5	5	1	Neutral
drp-1	12	20	0.6	Purifying
pdr-1	15	19	0.789	Purifying
eat-3	5	5	1	Neutral
fzo-1	5	5	1	Neutral

The dN/dS ratio measures the rate of non-synonymous mutations over the rate of synonymous mutations, and we used this to determine how genes in strains evolved by seeing which mutations were favored. Using aligned coding sequences from wild *C. elegans* strains, we found that *drp-1*, *pdr-1*, and *atfs-1* were under purifying selection, suggesting that changes in these genes were often deleterious and, in turn, selected for through evolution. Conversely, *pink-1*, *eat-3*, and *fzo-1* each had a dN/dS ratio of 1, indicating that these genes are evolving randomly.

Changes in Protein Structure Due to Missense Variants

Table 2. VEP results for candidate genes across CX11314 and EG4725.

Genes	Synonymous Variants	Missense Variants	Stop Gained
fzo-1	56%	44%	0
eat-3	52%	48%	0
drp-1	75%	25%	0
pink-1	59%	41%	0
atfs-1	55%	45%	0
pdr-1	53%	43%	3%

Table 3. List of specific amino acid substitutions found in the six genes of interest across CX11314 and EG4725 strains.

Gene	Number of Variants	Specific Variants
fzo-1	5	1 P->S, 1 R->Q, 1 A->T, 1 V->M, 1 L->I
eat-3	5	1 I->M, 1 A->P, 1 K->R, 1 V->A, 1 A->T
drp-1	12	4 K->R, 4 L->I, 4 T->I
pdr-1	15	3 S->T, 4 E->K, 4 G->D, 2 L->R, 2 F->L
atfs-1	34	7 P->L, 7 Q->R, 7 S->T, 6 V->A, 7 Y->N
pink-1	5	1 E->K, 1 A->S, 1 P->S, 1 V->M, 1 N->D

A Variant Effect Predictor (VEP) analysis on the 6 genes of interest for CX11314 and EG4725 then revealed notable missense variations, thereby affecting their protein structure. As shown in Figure 8, the variations made in *fzo-1* consisted of 56% synonymous variations, with the remaining 44% being missense variations. The conservative changes include Valine(nonpolar) to Methionine(nonpolar) and Leucine(nonpolar) to Isoleucine(nonpolar). On the other hand, nonpolar to polar amino acid substitution changes, such as Proline to Serine and Alanine to Threonine, as well as charge changes, such as a substitution from Arginine(polar positive) to Glutamine(polar neutral), introduce changes that will alter the structure and thereby function of the protein.

Secondly, the variations in *eat-3* consisted of 52% synonymous variation and 48% missense variation. In the case of eat-3, which consisted of 5 amino acid variations, most of the variations(Isoleucine to Methionine, Lysine to Arginine, and Valine to Alanine) were conservative. However, substitutions that change polarity, Threonine(polar) for Alanine(nonpolar), and changes in size, such as Alanine to Proline.

Even though *drp-1* had 12 missense variants, it had the highest synonymous variant percentages at 75%, suggesting that this gene has been evolutionarily conserved. In this case, *drp-1* had 4 of each type of missense variation: Lysine to Arginine, Leucine to Isoleucine, and Threonine to Isoleucine. All these, with the exception of Threonine to Isoleucine, are conservative changes that preserve its characteristic

Unlike the other genes, 53% of the variants were synonymous, 43% were missense, and 3% were stop-gain, meaning that they potentially led to a truncated protein sequence with altered functionality in *drp-1*. In the 15 missense substitutions, there were a total of 6 substitutions (4 Glycine to Aspartic acid and 2 Leucine to Arginine) that resulted in a polarity change.

Interestingly, *atfs-1* had the highest number of missense variants at 34. 45% of the variants were missense, and 53% were synonymous; however, none of the substitutions caused by the missense variants resulted in a change in polarity. Instead, there were 7 substitutions (Glutamine to Arginine) that changed from a neutral to basic pH; nonetheless, this will still result in an altered structure and function of the protein.

Finally, *pink-1* showed the highest missense variant percentage of the 6 genes at 59%, followed by 41% synonymous variants. Furthermore, most of the missense variants resulted in significant shifts: Glutamic acid to Lysine changed pH from acidic to basic, Asparagine to Aspartic Acid caused an acidic pH change, and Alanine to Serine, as well as Proline to Serine, caused a change in polarity in the amino acid.

DEG Analysis Reveals Statistical Significance of Certain Transcripts

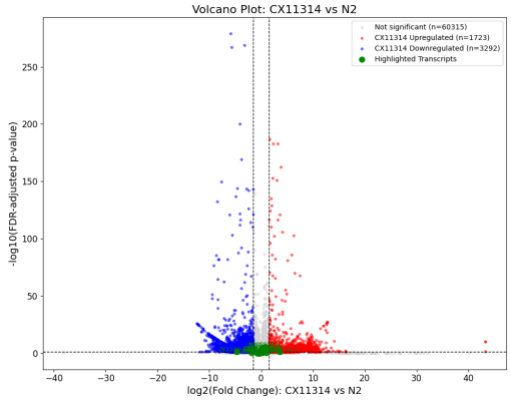

Figure 7. Volcano plots showing differential gene expression in CX11314 and N2.

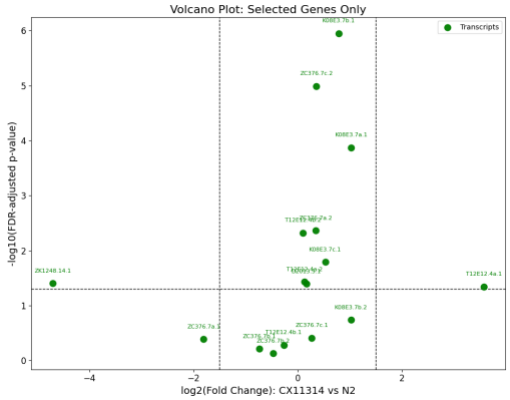

Figure 8. Zoomed-in view of selected mitochondrial genes of interest.

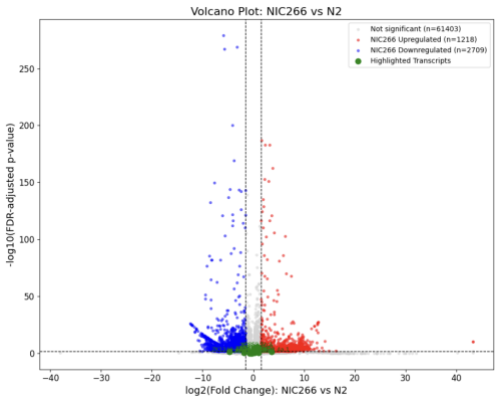

Figure 9. Volcano plots showing differential gene expression (log_2 fold change vs. p-value) between NIC266 and N2.

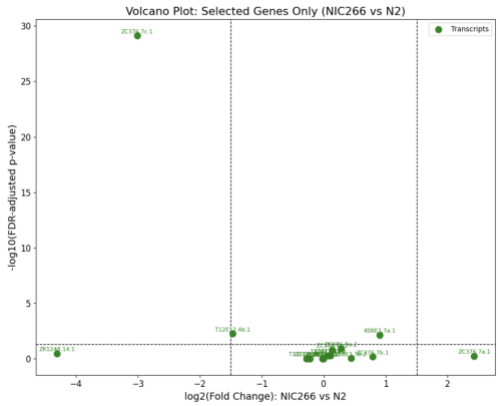

Figure 10. Zoomed-in view of selected mitochondrial genes of interest.

In bioinformatics, volcano plots are used to visualize how genes are differentially expressed between strains by plotting the statistical significance (p-value = 0.05) against the magnitude of expression change (log2 fold change). Figure 10a, which depicts the differential gene expression (DEG) between the wild strain CX11314 and N2, shows that most of the transcripts corresponding to the genes of interest didn't have statistical significance. For that matter, Figure 10b, which shows a zoomed-in visualization of the corresponding transcripts, illustrates how only 10 of the 16 transcripts were of statistical significance. Additionally, only two of the ten transcripts, *ZK1248.14.1* corresponding to *fzo-1*, and *T12E12.4a.1* corresponding to *drp-1*, were downregulated and upregulated, respectively. Although these genes are not highly regulated or of statistical significance, their consistent regulation suggests that they may, in fact, play crucial roles in fine-tuning mechanisms to ensure the longevity of the mitochondria. A similar scenario can be observed in Figure 11a. The volcano plot for the DEG of NIC266 vs N2 shows that only 3 of the 16 transcripts (*ZC276.7c.1* corresponding to *atfs-1, T12E12.4b.1* corresponding to *drp-1*, and *K083.7a.1* corresponding to *pdr-1*) were statistically significant, suggesting that there may be fewer transcriptional differences between NIC266 and N2. However, because NIC266 is used as a proxy for EG4725, further investigation would be needed to clarify the expression of the gene in EG4725.

Higher Expression of pdr-1, atfs-1, and eat-3

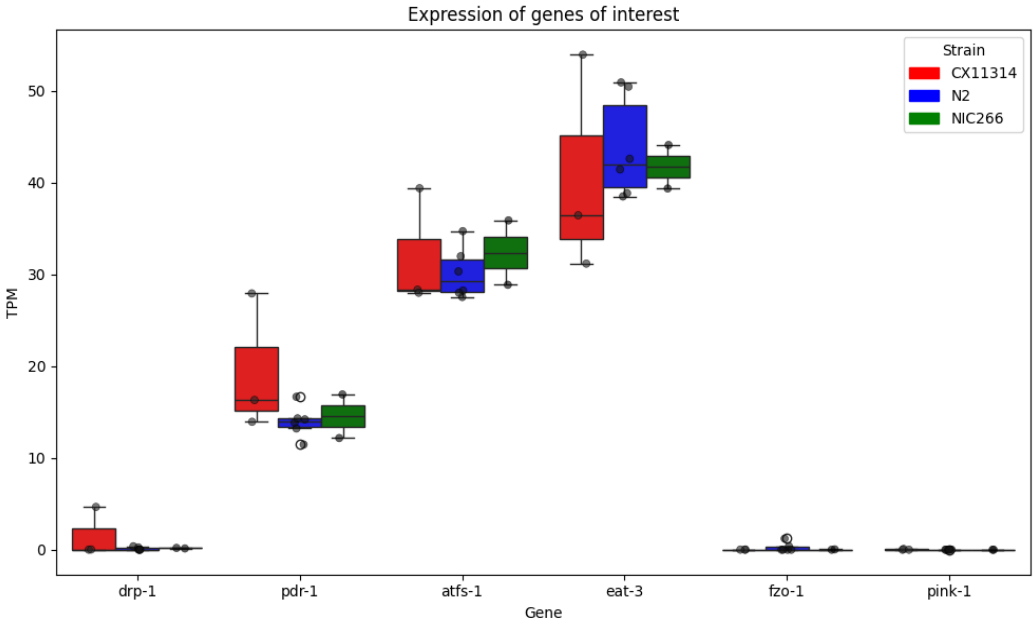

Figure 11. Box Plot comparing the expressions of genes in CX11314 and NIC266 with the common N2 strain, observing the Transcripts per Million(TPM)

The figure above depicts the TPM expression levels of the six mitochondrial genes of interest across N2, CX11314, and NIC266. While there is no apparent statistical significance between strains in the genes, *eat-3* had the highest overall expression among the six genes. Additionally, *pdr-1* had the highest expression in CX11314 compared to other strains, while *fzo-1* and *pink-1* had little to no expression in all the strains.

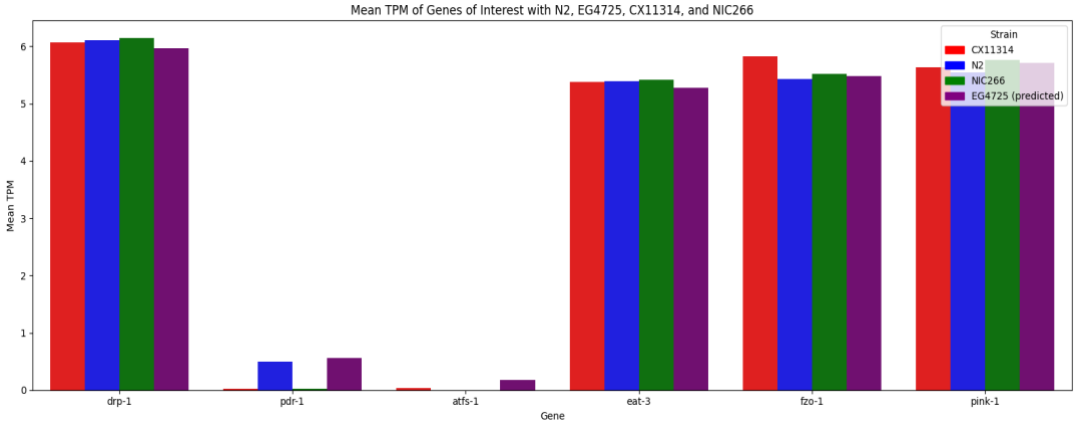

Figure 12. Bar Graph comparing the mean TPM of N2, CX11314, NIC266, and EG4725(predicted).

Figure 12 exhibits how the predicted expression profile for EG4725 across the six genes of interest aligns with that of N2, CX11314, and NIC266. This Machine Learning model was trained on the SNP and gene expression data of 204 *C. elegans* strains. Then, a Neural Network coupled with a Leave One Out Cross Validation (LOOCV) strategy, where NIC266 was left out of the training dataset to validate the model, was utilized. Finally, to quantify the predictiveness of our model, a Spearman Correlation was applied. This test revealed a prediction confidence of $\rho = 0.959$, indicating a strong correlation. This alignment demonstrates how implementing a machine learning approach to predict the expression of strains when there is a lack of data is a plausible technique. Moreover, with further refinement of the predictive model, it could be used to predict inaccessible gene expression data.

Discussion

After a comprehensive analysis, we were able to confidently conclude that natural genetic variation in these six key genes contributes to enhanced mitochondrial quality control in the wild *C. elegans* strains CX11314 and EG4725. Although the DEG analysis(Figures 10-12) suggests that there is no statistical significance in the respective gene expression levels across our strains of interest, we propose that these genes could potentially function by fine-tuning the processes for mitochondrial regulation to adapt to environmental stimuli. In fact, the PCA analysis (Figures 4-5) and phylogenetic tree (Figure 6) highlight the genetic divergence. Furthermore, the VEP analysis (Figure 8-9) also illustrated how the missense variants resulted in alterations in protein structure and function due to changes in polarity, charge, pH, and molecular size in the genes of CX11314 and EG4725. Interestingly, the gene expression prediction for EG4725 revealed higher expression of *pdr-1* and *atfs-1*, indicating that EG4725 could potentially have higher mitochondrial quality control. Overall, these findings support the notion that the natural genetic variations of genes in CX11314 and EG4725 contribute to their fitness in response to environmental pressures. Thus, correlating mitochondrial homeostasis genes in *C. elegans* with those in humans could prompt the use of therapeutic methods to combat mitochondrial diseases caused by impaired mitochondria in humans.

Therefore, we propose using CRISPR Cas9-mediated Homology Directed Repair (HDR) to swap both promoter and protein-coding regions of our genes of interest. This method is ideal since CRISPR Cas9 genome editing technology allows precise replacement of genomic regions with donor DNA sequences[25]. Specifically, we propose replacing the promoter and coding regions of N2 in *pdr-1* and *atfs-1* with those from EG4725. This is mainly due to the higher expression of *pdr-1* and *atfs-1* in EG4725, as witnessed in Figure 12. Figure 12 illustrates the expression of our genes of interest on a logarithmic scale, meaning that although there appears to be a slightly higher expression in EG4725, the actual value would be significantly higher. While there is little to no expression of *pdr-1* and *atfs-1* in the other strains, with the exception of *pdr-1* in N2, the predicted expression in EG4725 seems to be greater.

To appropriately conduct these experiments, we propose using four experimental groups: N2 (control), N2 with *pdr-1* allele from EG4725, N2 with *atfs-1* allele from EG4725, and EG4725 (baseline). This allows us to accurately compare the functional effects of individual gene variants from EG4725 within N2 to the native EG4725 strain. Which, in turn, will reveal insights into the phenotypic and behavioral transformations in N2. For example, we can perform phenotypic and behavioral assays to assess mitochondrial morphology, measure mitophagy-related activities, measure ATP and ROS production, and conduct stress resistance assays to analyze their effects on lifespan and overall activity. This will thereby help the scientific community understand how genetic variations drive phenotypic differences, as well as pave the way for novel therapeutic approaches to combat mitochondrial dysfunction in humans.

Discussion

This study investigated how natural genetic variation in mitochondrial health-regulating genes contributes to mitochondrial resilience in the Caenorhabditis elegans strains CX11314 and EG4725. By integrating variant analysis, gene expression information, and functional interpretation of key mitochondrial pathways, we identified consistent differences between these strains and the reference N2 strain. The results suggest that naturally occurring missense variants and regulatory differences in genes involved in mitochondrial dynamics, mitophagy, and stress response may collectively enhance mitochondrial stability and function. Our findings highlight the importance of studying natural genetic diversity, rather than relying solely on laboratory-induced mutations, to better understand how mitochondrial health is maintained under physiological conditions. The observed associations between genetic variation and improved mitochondrial resilience provide a useful framework for exploring how similar mechanisms may operate in other organisms. This work also demonstrates how combining genetic data with pathway-level interpretation can

generate biologically meaningful insights, even in relatively small datasets. While the present study focuses on *C. elegans*, the analytical approach may be extended to other model systems and complex traits related to cellular health and aging. Future studies could further validate these findings through targeted functional experiments and explore the broader implications of natural mitochondrial variation in stress tolerance and disease susceptibility.

Acknowledgements

We would like to thank Samantha Fiallo, Max Frank, and Ryan Son for their guidance in this process, as well as UCSB's Summer Research Academics for this opportunity.

Author Contribution Statement

Aaryan S., Mansi M., and Mohammed M. contributed equally to the writing of the paper, data collection, programming, graph and table generation, and data analysis. Generative AI tools, such as ChatGPT, were used to assist with Python programming, while Grammarly was employed to improve writing clarity and grammar.

References

1. Cleveland Clinic. Mitochondrial Diseases. Cleveland Clinic, May 9, 2023. https://my.clevelandclinic.org/health/diseases/15612-mitochondrial-diseases

2. Martikainen, M. H.; Majamaa, K. Incidence and Prevalence of mtDNA-Related Adult Mitochondrial Disease in Southwest Finland, 2009–2022: An Observational, Population-Based Study. BMJ Neurol. Open 2024, 6 (1), e000546. https://doi.org/10.1136/bmjno-2023-000546

3. Stańczyk, M.; Szubart, N.; Maślanka, R.; Zadrąg-Tęcza, R. Mitochondrial Dysfunctions: Genetic and Cellular Implications Revealed by Various Model Organisms. Genes 2024, 15 (9), 1153. https://doi.org/10.3390/genes15091153

4. Sterken, M. G.; Snoek, L. B.; Kammenga, J. E.; Andersen, E. C. The Laboratory Domestication of Caenorhabditis elegans. Trends Genet. 2015, 31 (5), 224–231. https://doi.org/10.1016/j.tig.2015.02.009

5. UCLA. UCLA Biologists Slow Aging, Extend Lifespan of Fruit Flies. https://newsroom.ucla.edu/releases/ucla-biologists-slow-aging-extend-lifespan-of-fruit-flies

6. Bratic, A.; Larsson, N.-G. The Role of Mitochondria in Aging. J. Clin. Invest. 2013, 123 (3), 951–957. https://doi.org/10.1172/JCI64125

7. Martins, A. C.; Virgolini, M. B.; Ávila, D. S.; et al. Mitochondria in the Spotlight: C. elegans as a Model Organism to Evaluate Xenobiotic-Induced Dysfunction. Cells 2023, 12 (17), 2124. https://doi.org/10.3390/cells12172124

8. Traa, A.; Keil, A.; Abdelrahman AlOkda; et al. Overexpression of Mitochondrial Fission or Mitochondrial Fusion Genes Enhances Resilience and Extends Longevity. Aging Cell 2024, e14262. https://doi.org/10.1111/acel.14262

9. Wallace, D. C. A Mitochondrial Paradigm of Metabolic and Degenerative Diseases, Aging, and Cancer: A Dawn for Evolutionary Medicine. Annu. Rev. Genet. 2005, 39, 359–407. https://doi.org/10.1146/annurev.genet.39.110304.095751

10. Zhang, G.; Roberto, N. M.; Lee, D.; Hahnel, S. R.; Andersen, E. C. The Impact of Species-Wide Gene Expression Variation on Caenorhabditis elegans Complex Traits. Nat. Commun. 2022, 13 (1), 1–14. https://doi.org/10.1038/s41467-022-31208-4

11. Yin, J. A.; Gao, G.; Liu, X. J.; et al. Genetic Variation in Glia–Neuron Signalling Modulates Ageing Rate. Nature 2017, 551 (7679), 198–203. https://doi.org/10.1038/nature24463

12. Zhu, Z.; Lu, Q.; Zeng, F.; Wang, J.; Huang, S. Compatibility between Mitochondrial and Nuclear Genomes Correlates with the Quantitative Trait of Lifespan in Caenorhabditis elegans. Sci. Rep. 2015, 5, 17303. https://doi.org/10.1038/srep17303

13. CaeNDR. Caenorhabditis elegans Natural Diversity Resource. 2023. https://caendr.org/

14. Miles, A.; Harding, N. scikit-allel: Explore and Analyse Genetic Variation, Version 1.3.3; 2021. https://scikit-allel.readthedocs.io/en/stable/ (accessed July 26, 2025).

15. Biopython Developers. Biopython. https://biopython.org/

16. Client Challenge. ete3. PyPI. 2025. https://pypi.org/project/ete3/ (accessed July 26, 2025).

17. PyDESeq2 Developers. PyDESeq2 Documentation, Version 0.4.12; 2023. https://pydeseq2.readthedocs.io/en/stable/

18. NumPy Developers. NumPy. 2024. https://numpy.org/

19. Matplotlib Developers. Matplotlib: Python Plotting — Matplotlib 3.1.1 Documentation. May 30, 2024. https://matplotlib.org/

20. Scikit-learn Developers. scikit-learn: Machine Learning in Python. 2024. https://scikit-learn.org/stable/

21. McLaren, W.; Gil, L.; Hunt, S. E.; et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016, 17 (1), 122. https://doi.org/10.1186/s13059-016-0974-4

22. NCBI GEO. GSE186719. NCBI | NLM | NIH. 2025. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE186719 (accessed July 26, 2025).

23. Statsig. Benjamini–Hochberg Procedure | Statsig Docs. October 14, 2024. https://docs.statsig.com/stats-engine/methodologies/benjamini%E2%80%93hochberg-procedure/

24. Spearman's Correlation. Statstutor, 2018. https://www.statstutor.ac.uk/resources/uploaded/spearmans.pdf

25. Liao, H.; Wu, J.; VanDusen, N. J.; Li, Y.; Zheng, Y. CRISPR/Cas9-Mediated Homology-Directed Repair for Precise Gene Editing. Mol. Ther.–Nucleic Acids 2024, 35, 102344. https://doi.org/10.1016/j.omtn.2024.102344

Authors

Aaryan Samanta. Aaryan Samanta, a sophomore at Legend College Preparatory in Cupertino, CA, intends to major in AI/CS. An AIME qualifier and AMC 10 & 12 national 1st-place perfect scorer, he earned dual USACO perfects in Gold and Silver. Lead author of an AIAM 2025 publication, he founded NextGenAI International and AI Ethos, Inc., advancing ethical, human-centered AI worldwide.

Mansi Mandal. Mansi Mandal, a junior at Mountain House High School, CA, is passionate about scientific discovery. Selected for the UCSB Summer Research Academy 2025, she explores molecular clock bioinformatics, integrating computational analysis with biology. Her academic journey reflects curiosity, dedication, and a drive to contribute meaningfully to future research.

Mohammed Yasar Meeran. Mohammed Meeran is a junior at Lebanon Trail High School passionate about computational biology and machine learning. He analyzes genomic datasets, studies gene expression patterns, and develops predictive models using biological data. He aspires to advance research bridging biology and computation with intended majors in Biology or Bioinformatics.